

Evaluation of Voice Codecs for the Australian Mobile Satellite System

Tony Bundrock, Mal Wilkinson,
Telecom Australia Research
Laboratories,
PO Box 249, Clayton, Vic.,
AUSTRALIA 3168
PHONE: +61 3 541 6421
FAX: +61 3 543 3339

ABSTRACT

This paper describes the evaluation procedure to choose a low bit rate voice coding algorithm for the Australian land mobile satellite system. The procedure is designed to assess both the inherent quality of the codec under 'normal' conditions and its robustness under various 'severe' conditions. For the assessment, 'normal' conditions were chosen to be a random bit error rate with added background acoustic noise and the 'severe' condition is designed to represent burst error conditions when the mobile satellite channel suffers fading due to roadside vegetation.

The assessment is divided into two phases. In the first phase, a reduced set of conditions is used to determine a short list of candidate codecs for more extensive testing in the second phase. The first phase conditions include quality and robustness and codecs are ranked with a 60:40 weighting on the two. In the second phase, the short-listed codecs are assessed over a range of input voice levels, BER's, background noise conditions, and burst error distributions. Assessment is by subjective rating on a five level opinion scale and all results are then used to derive a weighted Mean Opinion Score M_{total} using appropriate weights for each of the test conditions.

INTRODUCTION

The proposed Australian MOBILESAT system which is being developed by the Australian Government organisations, Aussat and Telecom Australia, is designed to provide circuit switched mobile voice/data services and packet data messaging services using the Aussat B series of satellites due to be launched in 1991/92. The economics and the practicality of this system dictate that the bit rate for the voice codec must be a minimum so that an individual voice channel will occupy minimum bandwidth and also require a minimum of satellite power. However, the quality of the voice channel must be adequate for interconnection to the PSTN and this must be maintained for the range of fading conditions expected on a mobile satellite channel.

There have been various low bit rate voice coding algorithms proposed for a mobile satellite system¹ but there has been no consensus on the best for this particular application. Therefore, to select the voice codec to be used for the MOBILESAT system, Aussat and Telecom Australia invited interested parties to submit codecs for an evaluation process to determine the one best suited to the specific requirements of a mobile satellite system. A document² providing details of the evaluation procedure was prepared and made available to all interested organisations. In late February of this year eight codecs were submitted to the Telecom Australia Research Laboratories, Melbourne, Australia, where the evaluation was conducted.

EVALUATION PROCEDURE

The voice codec evaluation and selection procedure is to be conducted in three phases. In the first two phases, a digital error generator test set is used to simulate the land mobile satellite channel. This test set enables evaluation of the inherent performance of the codec separate from the modem performance because it enables a baseband connection between coder and decoder. The test set is used to introduce random and burst error patterns. During the third phase, a QPSK modem and an analog mobile satellite channel simulator is used to carry out additional evaluation of the joint performance of the codec and modem. The QPSK modem also provides soft decision decoding outputs and differential encoding.

The first phase is designed for a quick evaluation of all codecs to determine a short list for a more extensive program in the second phase. During this first phase, speech quality and robustness is subjectively rated using a five point opinion scale. Quality is assessed by two conditions; (a) a random bit error rate of 10^{-3} , and, (b) an acoustic SNR of 15dB with zero channel error rate. Robustness is also assessed by 2 conditions; (a) a burst error pattern (described below), and, (b) a random bit error rate of 10^{-2} . A final rating is derived by determining a weighted average of the four conditions with a weighting of 60:40 in favour of the quality conditions.

The second phase assesses codecs over a wider range of conditions testing the effects of various random BERs, burst error patterns, voice input level and acoustic background noise. A final overall rating is then determined by using various weightings for each condition and the codecs are ranked accordingly. The codecs are also assessed for complexity, transmission delay, tone handling capability and tandem operation. Spot checks are also made on the performance with carbon microphones.

The source material for the evaluation consists of Harvard sentence pairs spoken by two male and two female English speaking talkers with clearly different voice

characteristics. The recordings are made with a linear microphone with spectrum weighting according to the Intermediate Reference System (IRS) specified in CCITT Recommendation P.48. All source material was normalised to the same level prior to playing through test codecs and all output is also adjusted to the same level so that a constant listening level is maintained.

The recorded test speech samples for the various codecs are randomised prior to the listening tests. The listening takes place in a quiet room using an Australian 200 series handset. The listening level is calibrated so that the average speech level of -10 dBPa is obtained with an IEC artificial ear. The assessment is based on a five point Absolute Category Rating (ACR) method with a scale of; (a) Excellent, (b) Good, (c) Fair, (d) Poor, and, (e) Bad, which are scored 5 to 1 for analysis purposes.

Reference conditions for the evaluation are obtained by using a Modulated Noise Reference Unit (MNRU) test unit. Eight MNRU dBQ values are used as reference conditions and these are randomised with the codec samples for the listening tests.

All test material was prepared using a computer controlled test fixture. The computer ensures all codecs are tested under identical conditions. For instance, the identical burst error pattern is applied for each codec and the passage of acoustic background noise is also identical. A standard PCM I/O interface was used between the codec and the source and test recordings.

BURST ERROR MODEL

In order to realistically test the codecs, a suitable channel model must be assumed which is able to reproduce the error phenomena expected to occur on the channel. Errors are produced on the channel by thermal noise when there is clear line of sight to the satellite and in this case the channel can be modelled as a memoryless Binary Symmetric Channel. In addition, the occurrence of roadside vegetation will attenuate the signal and produce error bursts. The length and severity of these bursts will depend on the fade duration and depth.

To reproduce the error phenomena in the codec test set a Markov model of the channel is used. In this model, the channel is described by a number of states each with a different BER. Transition probabilities are determined for each state and the channel may be viewed as one which switches back and forth between several BSCs, each with an associated BER. The Markov parameters were chosen from data which have been measured for various channels by the Telecom Australia Research Laboratories³. During the testing, burst error models for light and heavily shadowed channels are used.

INMARSAT

While Aussat and Telecom Australia were preparing for the evaluation and selection of a voice codec for the MOBILESAT system, Inmarsat were pursuing a very similar course for the Inmarsat-M system. Late in 1989 Inmarsat issued an RFT for an organisation to conduct an evaluation procedure of codec on its behalf. With a goal of producing a common standard for a codec for mobile satellite systems, Australia and Inmarsat discussed the two evaluation procedures and made changes to both so that comparisons can be readily made. Telecom Australia Research Laboratories also submitted a bid to conduct the Inmarsat evaluation and were awarded a contract for this purpose in January 1990. With some minor variations, the procedure described here is the same for the Inmarsat evaluation and the codecs submitted for evaluation are the same for both.

RESULTS

Eight codecs were submitted to the Telecom Australia Research Laboratories for evaluation. British Telecom also provided a codec using the 9.6 kbit/s algorithm chosen for the Inmarsat aeronautical service to provide a reference to the results obtained for the lower bit rate codecs. At this time, phase one of the evaluation has been completed and five codecs selected for the extensive phase two testing. The results are shown in Figures 1 and 2. Figure 1 shows the results obtained for a 10^{-3} BER averaged over all four speakers. The

error bars shown are the 95% confidence intervals. From Figure 1, we see a fairly close grouping with most codecs rating a MOS of better than 3 (Fair). For this test codec 3 ranks highest although, at the 95% confidence level, it is not statistically separate from codecs 1, 2, 4, 5 and 7. Figure 2 shows the results obtained when a burst error pattern corresponding to moderately severe shadowing (85% tree coverage) is applied. In this case, codec 3 is clearly superior. For the phase 2 evaluation, codecs 1 to 5 have been chosen. Figure 3 shows the MOS results obtained for the MNRU reference.

ACKNOWLEDGEMENTS

The contributions of the following are gratefully acknowledged: Mushfiq Rahman for the development of the Markov channel model and for planning the experimental sequencing; Alistair Urie for developing the computer controlled test fixture and his general contributions on codecs; Dan Cerchi, Adrian Martinus, and Steve Beyer for organising the source and test material; and Barry Thomas for conducting the subjective testing.

REFERENCES

1. Rafferty, W. 1988. Speech Compression Session, *Proceedings of the Mobile Satellite Conference*. JPL 88-9, (Pasadena, California; The Jet Propulsion Laboratory)
2. Wilkinson, M. H. 1990. *Voice Codec Test and Evaluation Procedure*. VCTEP/2 Telecom Research Laboratories, Australian Telecommunications Corporation, Melbourne Australia.
3. Bundrock, A. J., and Harvey, R. A. 1989 Propagation Measurements for an Australian Land Mobile Satellite System. *Australian Telecommunication Research*, Vol 23, No 1.

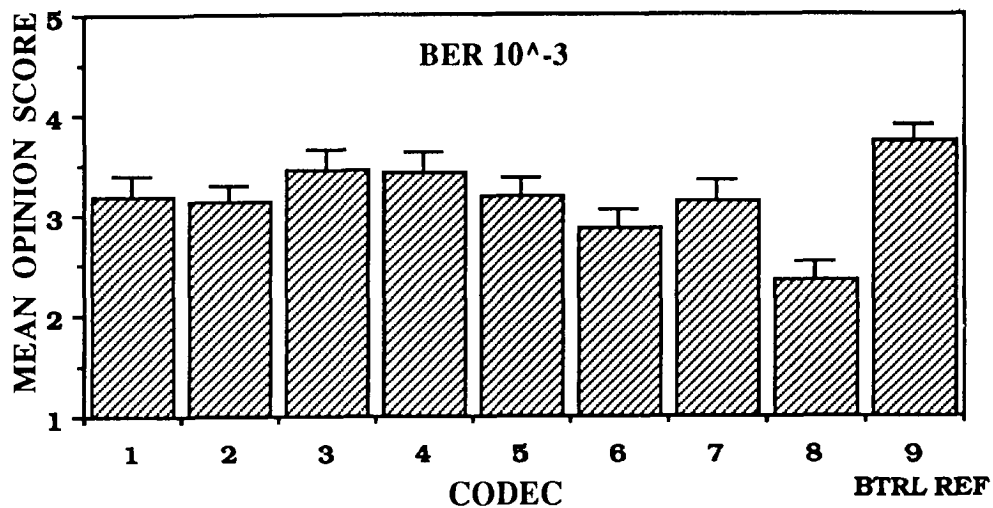


Fig. 1 Mean Opinion Score for CODECs

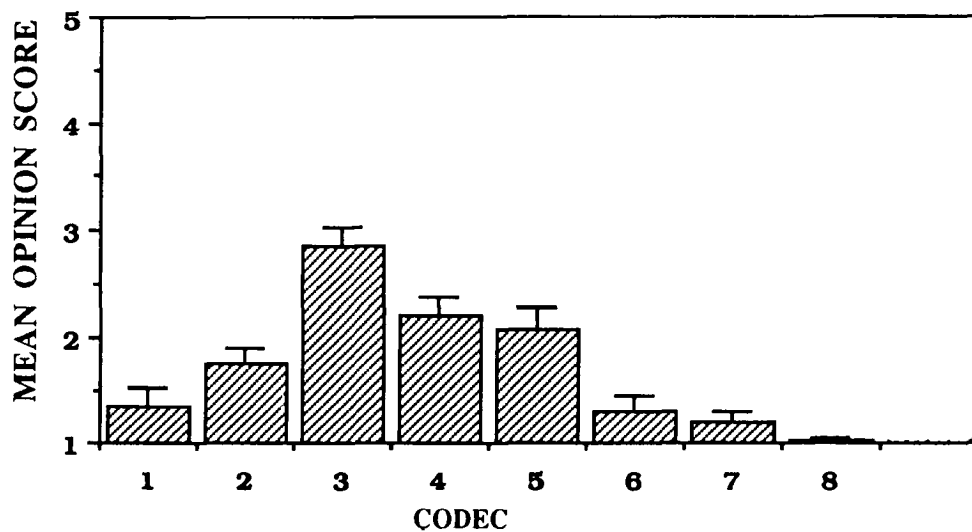


Fig. 2 Mean Opinion Score for Candidate CODECs with Burst Errors

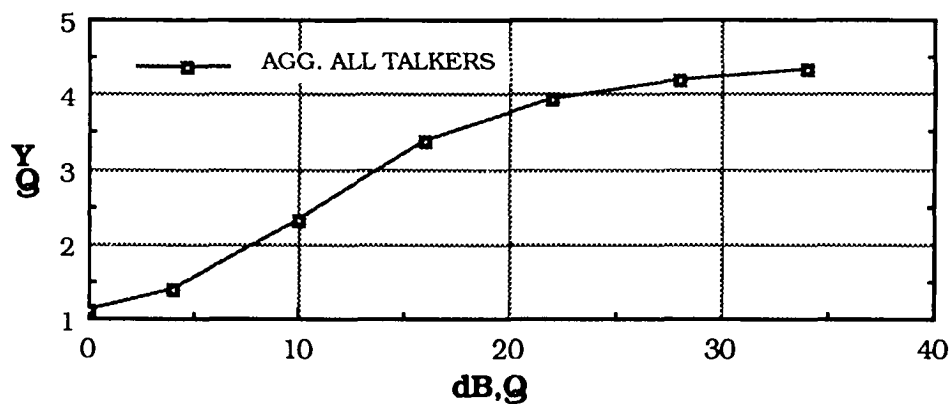


Fig. 3 MNRU Results for Test Listeners